

CLAIMS

1. A method for enhanced document retrieval, the method comprising:
receiving a search query from an end-user;
responsive to receiving the search query, retrieving search results, the search results comprising an enhanced document and a set of non-enhanced documents, the enhanced document and the non-enhanced documents including term(s) of the search query; and
wherein the enhanced document is derived from a base document, the base document having been modified with metadata mined from one or more different documents, the metadata being associated with one or more respective references to the base document, the one or more different documents being independent of the base document.
2. A method as recited in claim 1, wherein the set of non-enhanced documents is not an empty set.
3. A method as recited in claim 1, wherein term(s) of the search query pertain to a product research and/or product troubleshooting.
4. A method as recited in claim 1, wherein the base document is a knowledge base, product help, and/or developer targeted article.

5. A method as recited in claim 1, wherein the one or more different documents comprise one or more service request(s), newsgroup posting(s), and/or search query log(s).

6. A method as recited in claim 1, wherein the one or more respective references to the base document comprise a link and/or a substantially unique documents identifier associated with the base document.

7. A method as recited in claim 1, wherein the metadata comprises a title of a document, product problem context, and/or product problem resolution information.

8. A method as recited in claim 1, and further comprising:

determining relevance of the enhanced document and the set of non-enhanced documents in view of search query term proximity and popularity criteria; and

returning ranked search results to the end-user, the ranked search resulting being ranked as a function of the relevance.

9. A method as recited in claim 7, wherein the method further comprises calculating term proximity as follows:

$$Sim = Sim_{orig} * proximity,$$

$$proximity = \frac{\log(1 + \alpha(\beta * Hit + (1 - \beta) * (1 - EditDistance)))}{\log(1 + \alpha)},$$

wherein α , β are parameters configured to control relative weight of each part of the search query, Hit represents a percentage of the terms in a document in a database over all terms, the database comprising the one or more documents, and $EditDistance$ represents a misorder between the search query and the document.

10. A method as recited in claim 7, wherein the method further comprises calculating popularity as follows:

$$popularity = \frac{\log(1 + \alpha(\beta * I_{ref} + (1 - \beta) * (1 - I_{age})))}{\log(1 + \alpha)},$$

wherein α , β are parameters configured to control relative weight of each part of the search query, I_{ref} represents an importance from frequency of reference, and I_{age} represents an age of a document from a database associated with the base document, the document comprising at least a subset of the terms and/or keywords.

11. A method as recited in claim 10, wherein I_{ref} and I_{age} are determined as follows:

$$I_{ref} = 0.5 + 0.5 \frac{\text{freq(ref)}}{\text{max freq(ref)}}; \text{ and}$$

$$I_{age} = \frac{1}{1 + e^{age}}.$$

12. A method as recited in claim 7, wherein after determining the relevance and before returning the ranked results, the method further comprises:

creating a respective snippet description for each result of the top-ranked results, the snippet description indicating significance of the result in view of term(s) of the search query; and

wherein the ranked search results comprise the respective snippet description for each result of the top-ranked results.

13. A method as recited in claim 12, wherein creating further comprises:

locating one or more blocks from a retrieved document in the top-ranked search results; and

highlighting term(s) of the search query in the one or more blocks.

14. A method as recited in claim 13, wherein locating further comprises:

identifying the one or more blocks with a sliding window of configurable size that is applied to portions of the retrieved document;

measuring an amount of query-related information carried by text delineated by the sliding window, the measure being based on quantitative criteria such as word frequency, word proximity to a query term, and/or word position; and

combining the quantitative criteria with a trained classifier to identify a substantially most informative block for the snippet description.

15. A method as recited in claim 14, wherein the configurable size is a function of client computing device user interface space available for display of the snippet description.

16. A method as recited in claim 14, wherein the trained classifier is trained with linear regression as a function of:

$$y = b_0 + \sum_{j=1}^p b_j x_j + e$$

wherein x is a vector, y is a value of a straight line to fit value(s) associated with the quantitative criteria, "residual" e is a random variable with mean zero, coefficients b_j are determined by a condition that a sum of a square residual is small, variables x_j are inputs such as log or polynomial of inputs.

17. A computer-readable medium comprising computer-executable instructions providing content propagation for enhanced document retrieval, the computer-executable instructions comprising instructions for:

receiving a search query from an end-user;

responsive to receiving the search query, retrieving search results, the search results comprising an enhanced document and a set of non-enhanced documents, the enhanced document and the non-enhanced documents including term(s) of the search query; and

wherein the enhanced document is derived from a base document, the base document having been modified with metadata mined from one or more different documents, the metadata being associated with one or more respective references

to the base document, the one or more different documents being independent of the base document.

18. A computer-readable medium as recited in claim 17, wherein the set of non-enhanced documents is not an empty set.

19. A computer-readable medium as recited in claim 17, wherein the base document is a knowledge base, product help, and/or developer targeted article.

20. A computer-readable medium as recited in claim 17, wherein the one or more different documents comprise one or more service request(s), newsgroup posting(s), and/or search query log(s).

21. A computer-readable medium as recited in claim 17, wherein the one or more respective references to the base document comprise a link and/or a substantially unique documents identifier associated with the base document.

22. The computer-readable medium as recited in claim 17, wherein the metadata is semantically and/or contextually related to associated ones of the one or more documents.

23. A computer-readable medium as recited in claim 17, wherein the instructions further comprise instructions for:

determining relevance of the enhanced document and the set of non-enhanced documents in view of search query term proximity and popularity criteria; and

returning ranked search results to the end-user, the ranked search resulting being ranked as a function of the relevance.

24. A computer-readable medium as recited in claim 23, wherein the instructions further comprise instructions for calculating term proximity as follows:

$$Sim = Sim_{orig} * proximity,$$
$$proximity = \frac{\log(1 + \alpha(\beta * Hit + (1 - \beta) * (1 - EditDistance)))}{\log(1 + \alpha)},$$

wherein α , β are parameters configured to control relative weight of each part of the search query, *Hit* represents a percentage of the terms in a document in a database over all terms, the database comprising the one or more documents, and *EditDistance* represents a disorder between the search query and the document.

25. A computer-readable medium as recited in claim 23, wherein the instructions further comprise instructions for calculating popularity as follows:

$$popularity = \frac{\log(1 + \alpha(\beta * I_{ref} + (1 - \beta) * (1 - I_{age})))}{\log(1 + \alpha)},$$

wherein α , β are parameters configured to control relative weight of each part of the search query, I_{ref} represents an importance from frequency of reference, and I_{age} represents an age of a document from a database associated with the base document, the document comprising at least a subset of the terms and/or keywords.

26. A computer-readable medium as recited in claim 23, wherein after determining the relevance and before returning the ranked results, the instructions further comprise instructions for:

creating a respective snippet description for each result of the top-ranked results, the snippet description indicating significance of the result in view of term(s) of the search query; and

wherein the ranked search results comprise the respective snippet description for each result of the top-ranked results.

27. A computer-readable medium as recited in claim 26, wherein the instructions for creating further comprise instructions for:

locating one or more blocks from a retrieved document in the top-ranked search results; and

highlighting term(s) of the search query in the one or more blocks.

28. A computer-readable medium as recited in claim 27, wherein the instructions for locating further comprise instructions for:

identifying the one or more blocks with a sliding window of configurable size that is applied to portions of the retrieved document;

measuring an amount of query-related information carried by text delineated by the sliding window, the measure being based on quantitative criteria such as word frequency, word proximity to a query term, and/or word position; and

combining the quantitative criteria with a trained classifier to identify a substantially most informative block for the snippet description.

29. A computer-readable medium as recited in claim 28, wherein the configurable size is a function of client computing device user interface space available for display of the snippet description.

30. A computing device for enhanced document retrieval, the computing device comprising:

a processor; and

a memory coupled to the processor, the memory comprising computer-program instructions executable by the processor for:

receiving a search query from an end-user;

responsive to receiving the search query, retrieving search results, the search results comprising an enhanced document and a set of non-enhanced

documents, the enhanced document and the non-enhanced documents including term(s) of the search query; and

wherein the enhanced document is derived from a base document, the base document having been modified with metadata mined from one or more different documents, the metadata being associated with one or more respective references to the base document, the one or more different documents being independent of the base document.

31. A computing device as recited in claim 30, wherein the set of non-enhanced documents is not an empty set.

32. A computing device as recited in claim 30, wherein the base document is a knowledge base, product help, and/or developer targeted article.

33. A computing device as recited in claim 30, wherein the one or more different documents comprise one or more service request(s), newsgroup posting(s), and/or search query log(s).

34. A computing device as recited in claim 30, wherein the one or more respective references to the base document comprise a link and/or a substantially unique documents identifier associated with the base document.

35. A computing device as recited in claim 30, wherein the instructions further comprise instructions for:

determining relevance of the enhanced document and the set of non-enhanced documents in view of search query term proximity and popularity criteria; and

returning ranked search results to the end-user, the ranked search resulting being ranked as a function of the relevance.

36. A computing device as recited in claim 35, wherein the instructions further comprise instructions for calculating term proximity as follows:

$$Sim = Sim_{orig} * proximity,$$
$$proximity = \frac{\log(1 + \alpha(\beta * Hit + (1 - \beta) * (1 - EditDistance)))}{\log(1 + \alpha)}$$

wherein α , β are parameters configured to control relative weight of each part of the search query, *Hit* represents a percentage of the terms in a document in a database over all terms, the database comprising the one or more documents, and *EditDistance* represents a disorder between the search query and the document.

37. A computing device as recited in claim 35, wherein the instructions further comprise instructions for calculating popularity as follows:

$$popularity = \frac{\log(1 + \alpha(\beta * I_{ref} + (1 - \beta) * (1 - I_{age})))}{\log(1 + \alpha)},$$

wherein α , β are parameters configured to control relative weight of each part of the search query, I_{ref} represents an importance from frequency of reference, and I_{age} represents an age of a document from a database associated with the base document the document comprising at least a subset of the terms and/or keywords.

38. A computing device as recited in claim 35, wherein after determining the relevance and before returning the ranked results, the instructions further comprise instructions for:

creating a respective snippet description for each result of the top-ranked results, the snippet description indicating significance of the result in view of term(s) of the search query; and

wherein the ranked search results comprise the respective snippet description for each result of the top-ranked results.

39. A computing device as recited in claim 38, wherein the instructions for creating further comprise instructions for:

locating one or more blocks from a retrieved document in the top-ranked search results; and

highlighting term(s) of the search query in the one or more blocks.

40. A computing device as recited in claim 39, wherein the instructions for locating further comprise instructions for:

identifying the one or more blocks with a sliding window of configurable size that is applied to portions of the retrieved document;

measuring an amount of query-related information carried by text delineated by the sliding window, the measure being based on quantitative criteria such as word frequency, word proximity to a query term, and/or word position; and

combining the quantitative criteria with a trained classifier to identify a substantially most informative block for the snippet description.

41. A computing device for enhanced document retrieval, the computing device comprising:

receiving means to receive a search query from an end-user;

responsive to receiving the search query, retrieving means to retrieve search results, the search results comprising an enhanced document and a set of non-enhanced documents, the enhanced document and the non-enhanced documents including term(s) of the search query; and

wherein the enhanced document is derived from a base document, the base document having been modified with metadata mined from one or more different documents, the metadata being associated with one or more respective references to the base document, the one or more different documents being independent of the base document.

42. A computing device as recited in claim 30, wherein the set of non-enhanced documents is not an empty set.

43. A computing device as recited in claim 30, wherein the base document is a knowledge base, product help, and/or developer targeted article.

44. A computing device as recited in claim 30, wherein the one or more different documents comprise one or more service request(s), newsgroup posting(s), and/or search query log(s).

45. A computing device as recited in claim 30, wherein the one or more respective references to the base document comprise a link and/or a substantially unique documents identifier associated with the base document.

46. A computing device as recited in claim 30, wherein the computing device further comprises:

determining means to determine relevance of the enhanced document and the set of non-enhanced documents in view of search query term proximity and popularity criteria; and

returning means to return ranked search results to the end-user, the ranked search resulting being ranked as a function of the relevance.

47. A computing device as recited in claim 35, wherein the computing device further comprises:

creating means to create a respective snippet description for each result of the top-ranked results, the snippet description indicating significance of the result in view of term(s) of the search query; and

wherein the ranked search results comprise the respective snippet description for each result of the top-ranked results.

48. A computing device as recited in claim 38, wherein the creating means further comprises:

locating means to locate one or more blocks from a retrieved document in the top-ranked search results; and

highlighting means to highlight term(s) of the search query in the one or more blocks.

49. A computing device as recited in claim 39, wherein the locating means further comprise:

identifying means to identify the one or more blocks with a sliding window of configurable size that is applied to portions of the retrieved document;

measuring means to measure an amount of query-related information carried by text delineated by the sliding window, the measure being based on quantitative criteria such as word frequency, word proximity to a query term, and/or word position; and

combining the quantitative criteria with a trained classifier to identify a substantially most informative block for the snippet description.